

データの分析 (1) 箱ひげ図と四分位数・分散と標準偏差

<代表値>

- ① 平均値
- ② 中央値 (メジアン) ... データを大きさの順に並べたとき、中央にくる値
- ③ 最頻値 (モード) ... 最も度数が多いデータの値

<四分位数>

第1四分位数 Q_1 は下位の中央値。第3四分位数 Q_3 は上位の中央値。

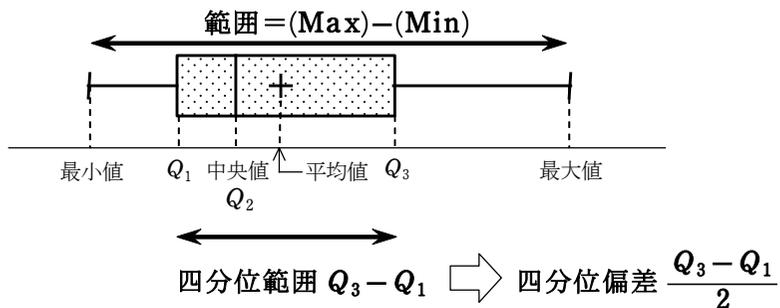
例 下位のデータ 上位のデータ

2	3	5	7	11	13	17	19	23	
		$Q_1=4$				$Q_2=11$			$Q_3=18$

例 下位のデータ 上位のデータ

2	3	5	7	11	13	17	19	23	29
		$Q_1=5$				$Q_2=12$			$Q_3=19$

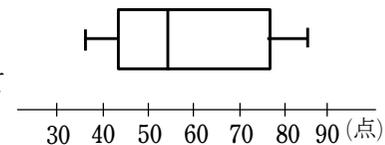
<箱ひげ図>



1 次のデータの第1四分位数, 第2四分位数, 第3四分位数を求めよ。

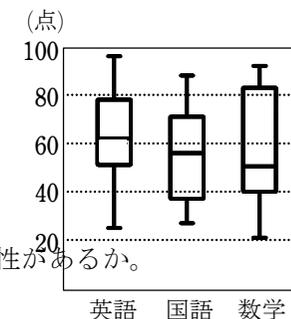
- (1) 10, 14, 15, 21, 28, 39, 53, 76, 99
- (2) 12, 35, 47, 59, 68, 73, 74, 79, 87, 97

2 右の図は, ある高校1年生 240 人に行った数学のテストについての, 得点の箱ひげ図である。この箱ひげ図から読み取れることとして正しいものを, 次の①~③ からすべて選べ。



- ① 30 点台の生徒はいない。
- ② 60 点未満の生徒は半数以上いる。
- ③ 40 点以上の生徒は 180 人以上いる。

3 右の図は, ある高校の1年生 50 人に行った英語, 国語, 数学のテストの得点を, 箱ひげ図に表したものである。



- (1) 80 点以上の生徒が 13 人以上いるのは, どの教科か。
- (2) 国語において, 60 点以下の生徒は最大で何人いる可能性があるか。また, 最小で何人いる可能性があるか。

4 次のデータは, ある本屋で1日あたりに売れた雑誌の冊数を8日間調べた結果である。ただし, a の値は 0 以上の整数である。

37 31 38 27 41 35 30 a (単位は冊)

a の値がわからないとき, このデータの中央値として何通りの値が考えられるか答えよ。

<分散と標準偏差>

①分散 $s^2 = \frac{1}{n} \{ \underbrace{(x_1 - \bar{x})^2}_{\text{平均から遠いデータが多い}} + \underbrace{(x_2 - \bar{x})^2}_{\text{平均に近いデータが多い}} + \dots + (x_n - \bar{x})^2 \}$

平均から遠いデータが多い \Rightarrow 分散 **大**

平均に近いデータが多い \Rightarrow 分散 **小**

②標準偏差 $s = \sqrt{\text{分散}}$

<分散公式 Part2> (分散) = (2乗の平均) - (平均)²

証明 $s^2 = \frac{1}{n} \{ (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \}$
 $= \frac{1}{n} \{ (x_1^2 + x_2^2 + \dots + x_n^2) - 2\bar{x}(x_1 + x_2 + \dots + x_n) + n(\bar{x})^2 \}$
 $= \frac{1}{n} (x_1^2 + x_2^2 + \dots + x_n^2) - 2\bar{x} \cdot \frac{1}{n} (x_1 + x_2 + \dots + x_n) + (\bar{x})^2$
 $= \overline{x^2} - 2\bar{x} \cdot \bar{x} + (\bar{x})^2$
 $= \overline{x^2} - (\bar{x})^2$

ただし、 $\overline{x^2}$ は、 x^2 のデータ $x_1^2, x_2^2, \dots, x_n^2$ の平均値を表す。

5 次の変数 x, y のデータは、X 高校、Y 高校に通う生徒 8 人ずつに小説を何冊持っているかを調査した結果である。

x 6, 3, 2, 10, 3, 1, 4, 11

y 7, 11, 6, 2, 10, 9, 12, 7 (単位は冊)

(1) 変数 x, y の平均値、分散、標準偏差を、それぞれ求めよ。ただし、小数第 2 位を四捨五入せよ。

(2) 変数 x, y について、標準偏差によってデータの平均値からの散らばりの度合いを比較せよ。

6 次のデータは、ある生徒 6 人について、けん垂が何回できたかを記録したものである。

14, 11, 10, 18, 16, 9 (単位は回)

(1) このデータの平均値を求めよ。

(2) このデータには一部に記録ミスがあり、正しくは 18 回が 17 回、9 回が 10 回であった。この誤りを修正したとき、データの平均値と分散は、修正前より増加するか、減少するか、変化しないかを答えよ。